CSCI 3901
(Fall 2022)
Assignment 5

External documentation:

1. **Overview:** I have connected the database to the java program. Summary information is extracted from the database, and I have stored the information in a file. The format of the file is XML file.

2. Classes used: I have used two classes for this assignment. The classes are:
    a. **Main Class**: This class is used to take three inputs from the user. The class takes the 'start' and 'end' date. It also asks user to input the file which will be generated in XML format. It also validates if the format of date is correct or not.
    b. XMLFile class: This class is connected to MySQL through JDBC**.** The XML file is generated using this class. It contains the functions responsible of generating the xml file.

3. Functions used:
    a. String sqlQuery(String start, String end)
        i. This method is used to connect the database to the java file. Using this method, we retrieve the data which we got in the sql queries.
    b. String printString(String fin, String start, string end)
        i. In this we return the string which contain the data in xml format.
    c. Writexml(String w, String file)
        i. This method is used to write the string in a file. This also generates a file in xml format.

4. Data structures used:
    a. Array List: I have used array list to store and retrieve the data which I got from sql queries.

5. Why my solution is ready to be deployed?
    a. I have made my queries easy to understand. If I want to add any attribute in my query it will be easy to add or remove attributes. New functionality such as new columns can be easily added.
    b. My code is divided into three functions sqlQuery, printString and Writexml to make my code more readable and maintainable.
    c. I have made a separate Main class to take the inputs from the user.
    d. I have used exceptional handling in my code to handle the errors which can occur during the execution of program.
    e. I have also tested my queries and their respective outputs and after the proper testing I have stored them in resultSet in my java code.
    f. I have used array list to efficiently store the data.
    g. I have used sqlQuery and printString functions to generate the string which is in XML format.

h. I have created writexml function which is used to generate a file in XML format. For this I have taken reference from w3Schools.com.
i. The code that I have written is robust, easy to understand and easy to maintain.

6. SQL Queries

a. Customer information

Report the customer name, address, value of orders, and owing balance for customers who became customers (have their first order) in the given time period.

QUERY

I have taken the new customer who had their first order between the between the provided dates. To calculate the outstanding balance I took the sum of the amount from the payment table and subtract it from the value of orders.

```
use csci3901;
SELECT
    a.customerNumber,a.customerName,a.addressLine1,a.city,a.state,a.postalCode,a.country,a.s AS value_of_orders,
    (SUM(p.amount) - a.s) AS outstanding_balance
FROM
    (SELECT
        c.customerNumber,c.customerName,o.orderNumber,c.addressLine1,c.city,c.state,c.postalCode,c.country,
            SUM(od.priceEach * quantityOrdered) AS s
    FROM
        orders o, customers c, orderdetails od
    WHERE
        o.customerNumber = c.customernumber
            AND o.orderNumber = od.orderNumber
            AND o.orderDate = (SELECT
                MIN(o1.orderDate)
            FROM
                orders o1
            WHERE
                o.customerNumber = o1.customerNumber)
            AND o.orderDate BETWEEN '2003-02-12' AND '2003-02-19'
    GROUP BY o.orderNumber) a,
    payments p
WHERE
    a.customerNumber = p.customerNumber
GROUP BY a.orderNumber;
```

OUTPUT

customerNumber	customerName	addressLine1	city	state	postalCode	country	value_of_orders	outstanding_balan...
278	Rovelli Gifts	Via Ludovico il Moro 22	Bergamo	NULL	24100	Italy	52151.81	75377.88

b. Product information

Report, for each product whose first sale is in the given period, the product name, product line name, the date the product was first sold and, for each customer who bought the product in the given period, the customer name and total number of units ordered by that customer.

QUERY

```sql
SELECT
    p.productName,
    p.productLine,
    o.orderDate AS introduction_date,
    c.customerName,
    od.quantityOrdered
FROM
    (SELECT
        o.orderNumber, o.orderDate, od.productCode, o.customerNumber
    FROM
        orders o, orderdetails od
    WHERE
        od.orderNumber = o.orderNumber
            AND orderDate BETWEEN '2003-02-12' AND '2004-02-19') AS a,
    orderdetails od,
    orders o,
    products p,
    customers c
WHERE
    od.productCode = a.productCode
        AND o.orderNumber = od.orderNumber
        AND a.productCode = p.productCode
        AND c.customerNumber = a.customerNumber
GROUP BY od.productCode;
```

OUTPUT

productName	productLine	introduction_da...	customerName	quantityOrdered
1980s Black Hawk Helicopter	Planes	2003-02-17	Osaka Souveniers Co.	36
1917 Grand Touring Sedan	Vintage Cars	2003-01-06	Signal Gift Stores	30
1948 Porsche 356-A Roadster	Classic Cars	2003-03-03	Mini Gifts Distributors Ltd.	38
1995 Honda Civic	Classic Cars	2003-03-10	Euro+ Shopping Channel	38
1998 Chrysler Plymouth Prowler	Classic Cars	2003-01-31	Dragon Souveniers, Ltd.	24

c. Office information

Report, for each city office, their city, territory, number of staff at the office, number of new customers in the given period, and the value of sales to new customers in the given period.

I have used 2 queries to get the required output. First I find the number of staff in each office using offices and employee table. And other details are calculated using query 2.

i. For city, territory and number of staff
    1. Query

```sql
SELECT
    o.city,
    o.territory,
    COUNT(DISTINCT (e.employeeNumber)) AS employee_count
FROM
    offices o,
    employees e
WHERE
    e.officeCode = o.officeCode
GROUP BY o.city;
```

2. Output

city	territory	employee_count
Boston	NA	2
London	EMEA	2
NYC	NA	2
Paris	EMEA	5
San Francisco	NA	6
Sydney	APAC	4
Tokyo	Japan	2

ii. For number of new customers in the given period, and the value of sales to new customers:

1. Query

```sql
SELECT
    a.city,
    a.territory,
    c.customerName,
    SUM(e.quantityOrdered * e.priceEach) AS customer_sales_value
FROM
    offices a,
    employees b,
    customers c,
    (select o.customerNumber, o.orderNumber,o.orderDate from orders  o
    where
    o.orderDate=(select min(o1.orderDate) from orders o1
    where o.customerNumber= o1.customerNumber
    and o1.orderDate BETWEEN '2003-02-12' AND '2003-02-19' group by o1.customerNumber)) d,
    orderdetails e
WHERE
    a.officeCode = b.officeCode
        AND b.employeeNumber = c.salesRepEmployeeNumber
        AND c.customerNumber = d.customerNumber
        AND d.orderNumber = e.orderNumber
GROUP BY a.city;
```

2. Output

	city	territory	customerName	customer_sales_value
▶	Paris	EMEA	Rovelli Gifts	52151.81